

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)

## For the greater good? Patient and public attitudes to use of medical free text data in research

Ford, Elizabeth<sup>1\*</sup>, Stockdale, Jessica<sup>1</sup>, Jackson, Richard<sup>2</sup>, and Cassell, Jackie<sup>1</sup><sup>1</sup>Brighton and Sussex Medical School<sup>2</sup>Kings College London

### Objectives

Electronic health records (EHRs) contain rich information for understanding health conditions and their treatment. A large proportion of clinical information in EHRs is stored in narrative free text. This text is currently under-utilised due to privacy concerns, as it is harder to remove patient identifiers from text than from structured data. Automated de-identification of clinical text is now possible using heuristic or machine-learning-based systems.

We conducted a review of the literature on patient and public understanding and attitudes towards the use of patients' medical data for research, particularly seeking views on free text. The aim was to inform and develop a governance framework for the de-identification and use of medical free text for research, and to instigate a wider discussion on the topic.

### Approach

Electronic health records (EHRs) contain rich information for understanding health conditions and their treatment. A large proportion of clinical information in EHRs is stored in narrative free text. This text is currently under-utilised due to privacy concerns, as it is harder to remove patient identifiers from text than from structured data. Automated de-identification of clinical text is now possible using heuristic or machine-learning-based systems.

We conducted a review of the literature on patient and public understanding and attitudes towards the use of patients' medical data for research, particularly seeking views on free text. The aim was to inform and develop a governance framework for the de-identification and use of medical free text for research, and to instigate a wider discussion on the topic.

### Results

Research participants were positively inclined towards information in records being used in research "for the greater good". However, no clear patterns by age, ethnicity, education level or SES emerged as to who was more favourable to data use.

Participants generally trusted health care professionals and public sector researchers with de-identified medical data, whereas government health agencies and commercial entities were not trusted. No explicitly feared harms associated with data use were articulated. However the general objections appeared to be a dislike of personal data being exploited for commercial gain, and a dislike of personal data being moved around and used without personal knowledge or consent.

Notably the use of EHR medical text for research did not emerge as a specific patient/public concern. De-identification was important to participants but text was not identified as a distinct privacy risk.

### Conclusion

This review demonstrates that transparency about data usage, and working "for the greater good" rather than financial gain, appear to be the most important public concerns to be addressed when using patients' medical data.

Governance frameworks for using EHRs must now be enhanced to provide for the use of medical text. This will involve informing both regulators and the public about the current capabilities of automated de-identification, and developing other assurances to safeguard patients' privacy.

\*Corresponding Author:

Email Address: [e.m.ford@bsms.ac.uk](mailto:e.m.ford@bsms.ac.uk) (E. Ford)